DOCUMENT RESUME

ED 190 143 IR 008 661

AUTHOR Wanger, Judith
TITLE Evaluation of the Online Search Process.
PUB DATE Mar 80
NOTE 10p.: Paper presented at the National Online Information Meeting (New York, NY, March 25-27, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Evaluation: *Information Retrieval: Information Seeking: *Online Systems: *Performance Factors: *Professional Continuing Education: Search Strategies: *Training
IDENTIFIERS National Library of Medicine MD

ABSTRACT

This study evaluated online searching performed on the National Library of Medicine system to identify implications for training and continuing education. Searchers were selected with formal and informal training, from three different types of organization, and with differing levels of experience. It was found that most searchers have clearly mastered the basics of searching, and that performance was unaffected by source of training, experience, organizational, or even subject educational background. The most consistent source of difference in performance was the difficulty of the search itself. It becomes necessary to determine characteristics of search requests to determine what makes a search easy or difficult, and to identify inherent limitations in databases and retrieval systems that hinder searching. The study concludes that searching is a highly individualized process, that there are good and bad searching practices, that training programs need revision to identify and illustrate workable search techniques, and that the user community needs to debate the issues of search techniques. (RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED190143

EVALUATION OF THE ONLINE SEARCH PROCESS

by

Judith Wanger

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Judith Wanger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR008661

EVALUATION OF THE ONLINE SEARCH PROCESS

Presented by Judith Wanger, Cuadra Associates, Inc.
National Online Conference, New York City, March 26, 1980

Good morning. I am very pleased to have this opportunity to share with you today some of our thoughts about the results of an exploratory study that we have recently completed for the U.S. National Library of Medicine. Cuadra Associates and its subcontractor, King Research, were awarded a contract to evaluate the online searching that is being done on the NLM system, with the purpose of identifying implications for training and continuing education.

A major research question that we were asked to address was whether there is a difference in performance among several different groups of searchers, primarily the three groups that are shown in this first chart: those who have been formally trained by NLM versus those who have been trained or learned in some other way; those from 3 different types of organizations; and those with differing levels of experience, defined in terms of the average number of searches conducted on NLM each month.

(1)

To provide some context, let me summarize briefly the key experimental design elements of the study, as shown in this next chart.

(2)

- We sampled a total of 216 searchers according to the characteristics mentioned earlier. The final number of study participants, due to dropouts and unusable returns, was 191.

- Each searcher was assigned to a cell--figuratively speaking, of course--in one of six replicates--you can visualize one replicate from Part B of this chart, which shows two searchers in each cell. Each group of 36 searchers in a replicate was assigned the same set of 3 searches, shown in Part C: one CANCERLIT patient-care search that we had classified as an easy search; one MEDLINE search for research or teaching, classified as being of medium difficulty; and one MEDLINE search for research or teaching, classified as hard.

In total, we obtained observations on 18 different search requests that had been drawn from the real world, and had a total of 535 usable searches across the study population.

In the time available, I'd like to provide you with a flavor of the major findings and share some general thoughts on the implications of the findings for training and continuing educations programs and materials.

The first finding I would like to report is that most searchers in this study have clearly mastered the basic mechanical skills of searching. To address this area, we designed three separate rating scales, as shown in the next chart: a system-use efficiency rating; an error rating; and a dubious-practice rating scale. Examples of the types of behaviors we looked for in each of the areas are also shown on this chart. We assigned to each incorrect, inefficient, or dubious behavior a specific number of points for

(3)

5.28

the first occurrence and, as appropriate, a number of points for subsequent occurrences. These points were deducted from a perfect starting score of 100 for each searcher, to arrive at the final percentage or rating. Overall, the searchers performed extremely well in each of these areas. Across the 18 searches, the range of average scores on each rating scale was from 78% to 100%.

We also looked at performance in terms of search output, using recall and precision ratios. As you know, recall refers to the extent to which the search uncovers all of the known relevant documents on a given topic in the file. Precision refers to the percentage of relevant items in the material delivered to the requester. Although traditional measurements, they have some limitations that must be recognized. They cannot be taken as absolute performance measurements for a number of reasons, but primarily because they are not associated with an end user's expectations. If I were to ask you whether a search with 100% recall and 50% precision was better or worse than a search with 50% recall and 100% precision, you would probably say: well, it depends on the search and the requester's needs.

The major statistical finding with these performance data was that differences in performance were not due to differences in the way searchers were trained (using distinctions made in this study), in their experience levels, their organizational settings, nor even in their subject/educational background. The most consistent source of difference in performance--again, as measured by recall and precision--was the search itself. This finding is illustrated in the next chart.

(4)

- Searches 1-6 (MEDLINE search of "medium" difficulty):

  --Average Recall (approximately 30 searchers): 8%-51%
  --Average Precision: 60%-79%

- Searches 7-12 (MEDLINE search of "hard" difficulty):

  --Average Recall: 8%-28%
  --Average Precision: 11%-93%

- Searches 13-18 (CANCERLIT search of "easy" difficulty):

  --Average Recall: 8%-70%
  --Average Precision: 17%-94%

- Overall (18 searches)

  --Average Recall=23%
  --Average Precision=67%

But it is the differences across searches we find most interesting. They suggest to us that it is important for us to begin identifying the characteristics of search requests that make them more or less difficult to fulfill well through an online search. And, we should also seek to characterize

the inherent limitations in databases and retrieval systems for achieving high performance levels for some search requests.

For example, in one of the CANCERLIT searches, 9 of the searchers achieved 100% recall, but the best precision score achieved by these searchers was 44%. After a preliminary review of their strategies and output in this natural-language database, we concluded that it is doubtful whether any of these searchers could have maintained that 100% recall level and, at the same time, improved their precision scores significantly. Therefore, to say that 44% is not a "good" score in this case is not being realistic.

Another implication from these findings is that we need to illustrate in our instructional programs differences in performance that can result from the tradeoff that one must frequently make between recall and precision. One training goal should be to help searchers understand how to manipulate the system so that they can achieve what is the most desirable tradeoff for a given search request. For example, one could give searchers an initial formulation for a search problem which they would enter online, and then instruct them to manipulate the system first to gain greater recall and then again to improve precision. This exercise should be followed by a discussion to compare the strategies and the bibliographies that result from each attempt in a highly structured version of DIALOG's ONTAP system.

In the course of the study, we also identified some specific searching behaviors and searcher attitudes that also have implications for training and continuing education. Among the behaviors that concerned us, two in particular are of major importance. One was the general lack of browsing done on the study searches, either to check the adequacy of a search formulation and/or to improve the search formulation. The second was the almost exclusive use of controlled vocabulary on MEDLINE by some searchers. These behaviors may be related, and they may be due to any number of factors, including cost, personnel time, or even a kind of routine or habit that has been developed, which says that I will prepare my searches offline, go online with my planned search, get the search down to a reasonable number of citations--by limiting to English language or recent years--and do my final print.

Before discussing the issue of vocabulary approach, let me begin by noting that NLM's MEDLINE vocabulary is deservedly one of the most respected controlled vocabularies in the business. However, on certain searches, we were led to wonder whether searchers were operating on some blind faith or making conscious decisions about when to use that vocabulary exclusively, and when to supplement the controlled terms with natural-language terms from the titles and abstracts. Although limited in our study of this area, we believe that further analysis of the study searches would provide some useful insights into this controversial area.

Regarding browsing online, I suspect that some folks will argue that it isn't necessary. That's probably true if you have done a search a number of times, and know from experience the terms you are using will produce the best results. But offline planning will not always be a legitimate substitute for online browsing. And, offline evaluation of the final printout generally cannot

substitute for online browsing and re-formulation, as needed. Yes, one can certainly cut out extraneous, irrelevant materials from a final printout to achieve greater precision, but one can't add to a final bibliography what one has failed to retrieve.

If some of the behaviors we observed on the printouts are really a reflection of cost concerns, then maybe we need to upgrade more rapidly to 1200-baud terminals--or maybe we will need to wait for changes in pricing structures. In the meantime, we may need to teach and to learn some alternative searching processes, such as the iterative search. Some searchers have already adopted the style of going online with an initial planned strategy, printing a number of titles and vocabulary terms online, and then logging off (after saving their searches). In this way, they can review what has happened and decide what should be done next online without the pressures of reacting online and feeling the taxi meter clicking.

Among the attitudes revealed, one was particularly fun. In a final comment, one searcher said: "I hated the second search--MEDLINE hard--there were no really relevant documents. I feel this should have been 'checked out' before the requests were sent out." Well, we had "checked out" all of the requests, and we knew that some were not neat and straightforward. But, that's the real world--which we were trying to simulate. In fact, that characteristic was one of the contributors to our classification of "hard" searches. The message here may be that our instructional programs should include some lessons on coping with "bad" search requests, which can't always be detected until you go online.

In conclusion, let me note that the study results dramatically emphasized for us what has long been hypothesized--the fact that searching is a highly individualized process. However, it has also led us to believe even more strongly that there are still both good and bad searching practices, as well as practices of dubious virtue. The time has now come for both suppliers and the user community to debate the issues of what is and is not good searching practice, and to begin including in training programs some of the pitfalls of searching that can yield inefficient searching or inadequate results.

To debate these issues, we will need to see more training and continuing education programs--in workshops and other formats--devoted to the illustration of sample searches conducted by experts, along with the resulting bibliographies, and a discussion of some of the alternative strategies that were discarded. We also need to see the increased use of sample searches that show what <u>not</u> to do and why. For the new user, the availability of these kinds of search samples will help them to synthesize the basic mechanics into an understanding of the whole online search process. For the veteran searchers, these sample searches may help to re-vitalize their thinking about the search process and to learn some new, more efficient behaviors.

We may not always agree now on what constitutes good and bad searching practices, but if suppliers and experienced searchers don't begin to articulate the differences and back them up with some formal study, we won't know where and why we disagree. I hope that the 1980's will become a period of such formal evaluation by both suppliers and users.

NLM STUDY: EVALUATION OF THE ONLINE SEARCH PROCESS

STRATIFICATION VARIABLES

TYPE OF TRAINING

Formal: trained by NLM (initial or advanced)

Informal: trained by a colleague, through MEDLEARN, through self instruction, library school, etc.

ORGANIZATION TYPE

Academic Institutions

Health Care Institutions/Agencies

Research-related (R&D) Organizations (including government agencies, commercial firms, research institutions)

NLM SEARCHING EXPERIENCE (usage)

Low: 0 to 5 searches per month

Medium: 6 to 15 searches per month

High: 16 or more searches per month

Figure 1

NLM STUDY: EVALUATION OF THE ONLINE SEARCH PROCESS

KEY ELEMENTS IN THE EXPERIMENTAL DESIGN

A. Sample Size: 216 searchers

B. A Replicate (36 searchers)

	NLM EXPERIENCE		
	Low	Medium	High
FORMALLY TRAINED			
Academic	2	2	2
Health Care	2	2	2
R&D	2	2	2
INFORMALLY TRAINED			
Academic	2	2	2
Health Care	2	2	2
R&D	2	2	2

C. Study Searches (18)

Search 1: MEDLINE
(Medium; Researching/Teaching)

Search 2: MEDLINE
(Hard; Researching/Teaching)

Search 3: CANCERLIT
(Easy; Patient Care)

Figure 2

NLM STUDY: EVALUATION OF THE ONLINE SEARCH PROCESS

SEARCH PROCESS PERFORMANCE MEASURES

SYSTEM-USE EFFICIENCY RATING

-- Same term or EXPLODE entered more than once
-- Incorrect logic, corrected
-- Early use of STRINGSEARCH
-- Thrashing
-- Hopscotching

ERROR RATING

-- Incorrect use of logic, not corrected
-- Incorrect entry of STRINGSEARCH, not corrected
-- Incorrect entry that generates a system error message

DUBIOUS PRACTICE RATING

-- No browsing of citations prior to final printing
-- ANDing of 3 or more concepts in one statement
-- Using PRINT to browse
-- Not ORing/NOTing statements to create final set
(or, failure to remove duplicates from separate formulations)

Figure 3

Summary of Output Measures, by Search

		Average No. of Retrieved Citations	Average Recall Scores	Average Precision Scores
MEDLINE (medium)	Search 1	4.3	22.6%	63.7%
	Search 2	6.3	12.4%	78.9%
	Search 3	19.8	50.9%	66.5%
	Search 4	25.1	14.1%	70.7%
	Search 5	5.3	8.3%	67.0%
	Search 6	3.0	8.2%	59.9%
MEDLINE (hard)	Search 7	16.2	15.9%	74.4%
	Search 8	15.2	14.6%	92.6%
	Search 9	5.9	12.5%	78.8%
	Search 10	5.0	8.4%	67.2%
	Search 11	14.7	28.4%	78.6%
	Search 12	11.6	9.7%	10.5%
CANCERLIT (easy)	Search 13	9.9	16.8%	16.5%
	Search 14	15.1	70.2%	38.1%
	Search 15	18.2	25.1%	63.2%
	Search 16	10.5	47.0%	49.9%
	Search 17	14.5	37.3%	54.0%
	Search 18	9.1	7.5%	94.3%